

Figure 1: Amino Acid and Nucleic Acid Sequences of *Hypocrea jecorina* Cel7A

1 GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle
CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCGGCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTGCACTCA ACAGACAGGC TCCGTGGTCA

101 ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
TCGACGCCAA CTGGCGCTGG ACTCAGCTA CGAACAGCAG CACGAACCTGC TACGATGGCA ACACITGGAG CTCGACCCCTA TGTCTGTGACA ACGAGACCTG

201 ·AlaIysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
CGCGAAGAAC TCCTGTCTGG ACGGTGCCGC CTACGGCTCC ACGTACGGAG TTACCACGAG CGGTAAACAGC CTCCTCCATTG GCCTTGTGCAC CCAGTCTGGC

301 GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCCTTCGAT GTTGATGTTTT

401 ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCACC AACACCCGCTG GCGCCAAAGTA

501 ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
CGGCACGGG TACTGTGACA GCCAGTGTCC CCGCATCTG AAGTTCTCA ATGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGGGAAC

601 ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
ACGGGCAATTG GAGGACACGG AAGCTGTGCTG TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCCTTACCCC CCACCCCTGC ACGACTGTGC

701 ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACCTTA CTCGATAAC AGATATGGCG GCACITGGCG TCCCGATGGC TGGCACTGGA ACCCATACCG

801 ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
CCTGGCAAC ACCAGCTTCT ACGGCCCTTG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGGCATC

901 AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

1001 ·GgluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACACGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCGATG GTTCTGGTCA TGAGTCTGTG

1101 ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCCAGC

1201 SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
TCCGGTGTCC CTGCTCAGT CGAATCTCAG TCTCCCAACG CCAAGGTAC CTTTCTCCAAC ATCAAGTTTCG GACCCATTGG CAGCACCCGC AACCCCTAGCG

1301 ·GglyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
CGCGCAACCC TCCCGGCGGA AACCCGCCCTG GCACCCACCAC CACCCGCCGC CCAGCCACTA CCACITGGAAG CTCCTCCGGA CCTACCCAGT CTCACTACGG

1401 ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaserGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
CCAGTGGGC GGTATTGGCT ACAGCGGCC CACGGTCTGC GCCAGCGGCA CAACCTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G

Figure 2A: Multiple alignment of the CBH1 homologous sequences.

	1	50
T reesei mat	(1)	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNC
H orientalis mat	(1)	QSACTLQETETHPSLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNC
H schweinitzii mat	(1)	QSACTLQETETHPSLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNC
T. koninlangbra mat	(1)	QSACTIQAEHPPLTWQKCSSGGGCTSQQTGSVVIDANNRWTHATNSTTNC
T. pseudokoningii mat	(1)	QSACTLQETETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNC
Consensus	(1)	QSACTLQETETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNC
	51	100
T reesei mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA
H orientalis mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTADSLSIGFVTQSA
H schweinitzii mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTADSLSIGFVTQSA
T. koninlangbra mat	(51)	YDGNTWSSSLCPDNESCAKNCCLDGAAYASTYGVTTADSLSIGFVTQSQ
T. pseudokoningii mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTADSLSIGFVTQSA
Consensus	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTADSLSIGFVTQSA
	101	150
T reesei mat	(101)	QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSM
H orientalis mat	(101)	QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSM
H schweinitzii mat	(101)	QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSM
T. koninlangbra mat	(101)	QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSM
T. pseudokoningii mat	(101)	QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSM
Consensus	(101)	QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSM
	151	200
T reesei mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSNNAN
H orientalis mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSNNAN
H schweinitzii mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSNNAN
T. koninlangbra mat	(151)	ADGGVSKYPSNTAGAKYGTGYCDSQCPRDLKFINGEANVEGWEPFASNNAN
T. pseudokoningii mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGEANVEGWEPFASNNAN
Consensus	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSNNAN
	201	250
T reesei mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEI CDGDCGGGTYSND
H orientalis mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEI CDGDCGGGTYSND
H schweinitzii mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEI CDGDCGGGTYSND
T. koninlangbra mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQAI CDGDCGGGTYSDD
T. pseudokoningii mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEI CDGDCGGGTYSGD
Consensus	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEI CDGDCGGGTYS D
	251	300
T reesei mat	(251)	RYGGTCDDPGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSGAI
H orientalis mat	(251)	RYGGTCDDPGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSGAI
H schweinitzii mat	(251)	RYGGTCDDPGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSGAI
T. koninlangbra mat	(251)	RYGGTCDDPGCDWNPYRLGNTSXYGPSSFTLDTTKKMTVVTFQFATSGAI
T. pseudokoningii mat	(251)	RYGGTCDDPGCDWNPYRLGNTSFYGPSSFALDTTKKLTVVTFQFETSGAI
Consensus	(251)	RYGGTCDDPGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSGAI
	301	350
T reesei mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGGSSFSKDGGLT
H orientalis mat	(301)	NRYYVQNGVTYQQPNAELGSYSGNELNDDYCTAESEFGGSSFSKDGGLT
H schweinitzii mat	(301)	NRYYVQNGVTYQQPNAELGSYSGNELNDAYCTAESEFGGSSFSKDGGLT
T. koninlangbra mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNTLNDAYCAAEAEFGGSSFSKDGGLT
T. pseudokoningii mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNSLDDDYCAAEAEFGGSSFSKDGGLT
Consensus	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGGSSFSKDGGLT

Figure 2B: Multiple alignment of the CBH1 homologous sequences.

		351	400
T reesei mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
H orientalis mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
H schweinitzii mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
T. koninlangbra mat	(351)	QFKQATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAARGSCSTS	
T. pseudokoningii mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
Consensus	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
		401	450
T reesei mat	(401)	SGVPAQVESQSPNAKVTFSTNIKFGPIGSTGNPSGGNPPGG-NPPGTTTTTR	
H orientalis mat	(401)	SGVPAQLESQSPNAKVVSSTNIKFGPIGSTGNPSGGNPPGG-NPPGTTTTTR	
H schweinitzii mat	(401)	SGVPAQLESQSANAKVVYSSTNIKFGPIGSTGNPSGGNPPGG-NPPGTTTTTR	
T. koninlangbra mat	(401)	SGVPAQLESQSTNAKVVSSTNIKFGPIGSTGNSSGGNPPGGNPPGTTTTTR	
T. pseudokoningii mat	(401)	SGVPAQLESQSSNAKVVSSTNIKFGPIGSTGNSSGGSPGGNPPGTTTTTR	
Consensus	(401)	SGVPAQLESQS NAKVVYSSTNIKFGPIGSTGNPSGGNPPGG NPPGTTTTTR	
		451	499
T reesei mat	(450)	RPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTTCQ-VLNPPYYSQCL	
H orientalis mat	(450)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGTTTCQ-VLNPPYYSQCL	
H schweinitzii mat	(450)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTICASGTTTCQVLNEYYSQCL	
T. koninlangbra mat	(451)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
T. pseudokoningii mat	(451)	RPATSTGSSPGPTQTHYGQCGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
Consensus	(451)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGTTTCQ VLNEYYSQCL	

CGTCATCTCG	GCCTTCTTGG	CCACGGCCCCG	TGCTCAGTCG	GCCTGCACTC	50
TCCAAACGGA	GACTCACCCG	TCTCTGACAT	GGCAGAAATG	CTCGTCTGGC	100
GGCACTTGCA	CCCAGCAGAC	AGGCTCCGTG	GTCATCGACG	CCAAGTGGCG	150
CTGGACTCAC	GCGACTAACA	GCAGCACGAA	CTGCTACGAC	GGCAACACTT	200
GGAGCTCAAC	CCTATGCCCT	GACAACGAGA	CTTGCGCGAA	GAATTGCTGC	250
CTGGACGGTG	CCGCCTATGC	GTCCACGTAC	GGAGTCACCA	CGAGTGCCGA	300
CAGCCTCTCC	ATCGGCTTCG	TCACGCAATC	TGCACAGAAG	AACGTTGGCG	350
CCCGTCTCTA	CCTGATGGCG	AGTGACACGA	CTTACCAGGA	GTTACGCTG	400
CTTGGAACG	AGTTCTCTTT	TGACGTTGAT	GTTTCGCAGC	TGCCGTAAGT	450
GACAACCATT	CCCCGCGAGG	CCATCTTCTC	ATTGGTTCCG	AGCTGACCCG	500
CCGATCTAAG	ATGTGGCTTG	AACGGCGCTC	TGTACTTCGT	GTCTATGGAT	550
GCGGATGGTG	GCGTGAGCAA	GTATCCCACC	AACACCGCCG	GCGCCAAGTA	600
CGGCACGGGC	TACTGCGACA	GCCAGTGCCC	CCGCGATCTC	AAGTTCATCA	650
ACGGCCAGGC	CAACGTTGAA	GGCTGGGAGC	CGTCTCCAA	CAACGCCAAC	700
ACGGGTATTG	GCGGACACGG	AAGCTGCTGC	TCTGAGATGG	ATATCTGGGA	750
GGCCAACTCC	ATCTCCGAGG	CTCTGACTCC	TCACCCTTGC	ACGACTGTTG	800
GCCAGGAGAT	CTGCGACGGT	GACGGCTGCG	GCGGAACCTA	CTCCAACGAC	850
CGATATGGTG	GTAATTGCGA	TCCTGATGGT	TGTGATTGGA	ATCCATACCG	900
CTTGGGCAAC	ACCAGCTTCT	ATGGCCCTGG	CTCGACCTTC	ACCCTCGATA	950
CCACCAAGAA	GTTGACCGTT	GTCAACCCAGT	TCGAGACCTC	GGGTGCCATC	1000
AACCGTTACT	ATGTCCAGAA	CGGCGTCACT	TACCAGCAAC	CCAACGCCGA	1050
GCTCGGTAGT	TACTCTGGTA	ATGAGCTCAA	CGATGACTAC	TGCACAGCTG	1100
AGGAGTCGGA	ATTTCGGCGGC	TCCTCCTTCT	CGGACAAGGG	CGGCCTTACT	1150
CAGTTCAAGA	AGGCCACTTC	CGGCGGCATG	GTCCTGGTCA	TGAGCTTGTG	1200
GGATGACGTG	AGTTGATAGA	CAGCATTAC	ATTGTCGTTG	GAAAGACGGG	1250
CGGCTAACCG	AGACATATGA	TATCTAACAG	TACTACGCCA	ACATGCTGTG	1300
GCTGGACTCC	ACCTACCCGA	CAAACGAGAC	CTCCTCCACC	CCCGGCGCCG	1350
TGCGCGGAAG	CTGCTCCACC	AGCTCCGGCG	TCCCCGCTCA	GCTCGAGTCC	1400
CAGTCCCCCA	ACGCCAAGGT	CGTCTACTCC	AACATCAAGT	TCGGGCCCCAT	1450
TGGCAGCACC	GGCAACCCCA	GCGGCGGAAA	CCCTCCTGGC	GGAAACCCCTC	1500
CCGGCACCAC	CACCACCCGC	CGCCCAGCTA	CCACCACTGG	AAGCTCTCCC	1550
GGACCTACTC	AGACTCACTA	CGGCCAGTGC	GGCGGCATCG	GCTACAGCGG	1600
CCCTACGGTC	TGCGCCAGCG	GCACGACCTG	CCAGG		1635

Figure 3: *H. oreintalis* genomic DNA sequence.

MYRKLAVISA FLATARA

17

Figure 4A: H. orientalis amino acid signal sequence.

QSACTLQTET	HPSLTWQKCS	SGGTCTQQTG	SVVIDANWRW	THATNSSTNC	50
YDGNTWSSTL	CPDNETCAKN	CCLDGAAYAS	TYGVTTADS	LSIGFVTQSA	100
QKNVGARLYL	MASDTTYQEF	TLLGNEFSFD	VDVSQLPCGL	NGALYFVSMD	150
ADGGVSKYPT	NTAGAKYGTG	YCDSQCPRDL	KFINGQANVE	GWEPSSNNAN	200
TGIGGHGSCC	SEMDIWEANS	ISEALTPHPC	TTVGQEICDG	DGCGGTYSND	250
RYGGTCDPDG	CDWNPYRLGN	TSFYGPGSSF	TLDTTKKLTV	VTQFETSGAI	300
NRYYVQNGVT	YQQPNAELGS	YSGNELNDDY	CTAESEFFGG	SSFSDKGGLT	350
QFKKATSGGM	VLVMSLWDDY	YANMLWLDST	YPTNETSSTP	GAVRGSCSTS	400
SGVPAQLESQ	SPNAKVVSYN	IKFGPIGSTG	NPSGGNPPGG	NPPGTTTTRR	450
PATTTGSSPG	PTQTHYGQCG	GIGYSGPTVC	ASGTTCQVLN	PYYSQCL	497

Figure 4B: H. orientalis mature amino acid sequence.

TCGGCCTGCA	CTCTCCAAAC	GGAGACTCAC	CCGTCTCTGA	CATGGCAGAA	50
ATGCTCGTCT	GGCGGCACCT	GCACCCAGCA	GACAGGCTCC	GTGGTCATCG	100
ACGCCAACTG	GCGCTGGACT	CACGCTACTA	ACAGCAGCAC	GAAGTGCTAC	150
GACGGCAACA	CTTGGAGCTC	AACCCGTGTG	CCTGACAATG	AGACTTGCGC	200
GAAGAACTGC	TGCCTGGACG	GTGCCGCCTA	TGCGTCCACG	TACGGAGTCA	250
CCACGAGTGC	CGACAGCCTC	TCCATCGGCT	TCGTGACACA	GTCTGCACAG	300
AAAAACGTTG	GCGCCCGTCT	CTACCTGATG	GCGAGTGACA	CGACTTACCA	350
GGAGTTCACG	CTGCTTGGA	ACGAGTTCTC	ATTCGACGTT	GATGTTTCGC	400
AGCTGCCGTA	AGTGACAACC	ATTCCCCCGA	CGCCATCTTC	TCATTGGTTC	450
GAAGCTGACC	CGCCGATCTA	AGATGTGGCT	TGAACGGCGC	TCTTTACTTC	500
GTGTCCATGG	ACGCAGATGG	TGGCGTGAGC	AAGTATCCCA	CCAACACCGC	550
CGGCGCCAAG	TACGGCACGG	GCTACTGTGA	CAGCCAGTGC	CCCCGCGATC	600
TCAAGTTTAT	CAACGGCCAG	GCCAACGTTG	AAGGCTGGGA	GCCGTCCTCC	650
AACAACGCCA	ACACGGGTAT	TGGCGGACAC	GGAAGCTGCT	GCTCCGAGAT	700
GGATATCTGG	GAGGCCAACT	CCATCTCCGA	GGCTCTTACT	CCTCACCTTT	750
GCACGAATGT	TGGCCAGGAG	ATCTGCGACG	GTGACGGCTG	CGGCGGAACC	800
TACTCCAACG	ACCGATATGG	TGGTACTTGC	GATCCTGATG	GTTGTGATTG	850
GAATCCATAC	CGCTTGGGCA	ACACCAGCTT	CTATGGCCCT	GGCTCGAGCT	900
TCACCCTCGA	TACCACCAAG	AAGTTGACCG	TCGTCACCCA	GTTTCGAGACT	950
TCGGGTGCCA	TCAACCGTTA	CTATGTCCAG	AATGGCGTCA	CTTACCAGCA	1000
ACCCAACGCC	GAGCTCGGCA	GTTACTCTGG	TAATGAGCTC	AACGATGCCT	1050
ACTGCACAGC	TGAAGAGTCG	GAATTTGGCG	GTTCTCTCTT	CTCGGACAAG	1100
GGCGGCCTTA	CTCAGTTCAA	GAAGGCCACT	TCCGGCGGCA	TGGTCCTGGT	1150
CATGAGCTTG	TGGGATGACG	TGAGTCCATA	GAACAGCATT	CACATTGTCTG	1200
TCGGAAAGAC	GGCCGGCTAA	CCGAGACATT	ACAGTACTAC	GCCAACATGC	1250
TGTGGCTGGA	CTCCACCTAC	CCGACAAACG	AGACCTCCTC	CACCCCGGT	1300
GCCGTGCGCG	GAAGCTGCTC	CACCAGCTCC	GGCGTCCCAG	CTCAGCTCGA	1350
GTCCCAGTCC	GCCAACGCCA	AGGTCGTCTA	CTCCAACATC	AAGTTCGGAC	1400
CCATTGGCAG	CACCGGCAAC	CCCAGCGGCG	GAAACCCTCC	TGGCGGAAAC	1450
CCTCCCGGCA	CCACCACCAC	CCGCCGCCCA	GCTACCACCA	CTGGAAGCTC	1500
TCCCGGACCT	ACTCAGACTC	ACTATGGCCA	GTGCGGCGGC	ATCGGCTACA	1550
GCGGCCCTAC	GATCTGCGCC	AGCGGCACGA	CCTGCCAGG		1589

Figure 5: *H. scweinitzii* genomic DNA sequence.

MYRKLAVITA FLATARA

17

Figure 6A: H. Schweinitzii signal peptide.

QSACTLQTET	HPSLTWQKCS	SGGTCTQQTG	SVVIDANWRW	THATNSSTNC	50
YDGNTWSSTL	CPDNETCAKN	CCLDGAAYAS	TYGVTTADS	LSIGFVTQSA	100
QKNVGARLYL	MASDTTYQEF	TLLGNEFSFD	VDVSQLPCGL	NGALYFVSMD	150
ADGGVSKYPT	NTAGAKYGTG	YCDSQCPRDL	KFINGQANVE	GWEPSSNNAN	200
TGIGGHGSCC	SEMDIWEANS	ISEALTPHPC	TNKGQEICDG	DGCGGTYSND	250
RYGGTCDDPG	CDWNPYRLGN	TSFYGPGSSF	TLDTTKKLT	VTQFETSGAI	300
NRYYVQNGVT	YQQPNAELGS	YSGNELNDAY	CTAESEFEGG	SSFSDKGGLT	350
QFKKATSGGM	VLVMSLWDDY	YANMLWLDST	YPTNETSSTP	GAVRGSCSTS	400
SGVPAQLESQ	SANAKVVYSN	IKFGPIGSTG	NPSGGNPPGG	NPPGTTTTRR	450
PATTTGSSPG	PTQTHYGQCG	GIGYSGPTIC	ASGTTQVLN	PYYSQCL	497

Figure 6B: H. Schweinitzii mature amino acid sequence. 497 residues

TCGGCCTGCA	CCATTCAAGC	GGAGACTCAC	CCGCCTCTGA	CATGGCAGAA	50
ATGCTCATCC	GGTGGTAGTT	GCACCTCGCA	AACCGGTTCT	GTGGTGATTG	100
ACGCGAAGCTG	GCGATGGACT	CACGCGACTA	ACAGCACCAC	GAAGTGCTAC	150
GACGGTAACA	CTTGGAGCTC	CAGTCTTTGC	CCCGACAATG	AGAGTTGCGC	200
AAAGAACTGC	TGCCTGGACG	GTGCAGCCTA	CGCATCCACG	TACGGAGTCA	250
CCACGAGTGC	TGATAGCCTC	TCCATTGGCT	TCGTCACTCA	GTCTCAGCAG	300
AAGAATGTTG	GCGCTCGTCT	CTACCTGATG	GCAAGCGACA	CGACCTACCA	350
GGAATTTACC	CTGCTTGGCA	ACGAGTTCTC	TTTCGATGTT	GATGTTTCAC	400
AGCTGCCGTA	AGTGACTAGC	ATTTACCTCC	GACGCCATCT	CATTGATTCC	450
CAGCTGACGG	CCAATTCAAG	ATGTGGCTTG	AACGGAGCCC	TTTACTTCGT	500
GTCCATGGAC	GCGGATGGTG	GCGTGAGCAA	GTATCCCTCC	AACACTGCCG	550
GCGCCAAGTA	CGGCACGGGC	TACTGCGATA	GCCAGTGTC	CCGTGATTTG	600
AAGTTTCATCA	ACGGCGAGGC	CAACGTTGAG	GGCTGGGAGC	CGGCTTCGAA	650
CAACGCCAAC	ACGGGTATTG	GCGGACACGG	AAGCTGCTGC	TCTGAGATGG	700
ATATCTGGGA	GGCCAACTCC	ATCTCTGAGG	CCCTTACTCC	TCACCCCTGC	750
ACGACTGTCTG	GCCAGGCCAT	TTGCGATGGT	GACGGCTGCG	GTGGAACCTA	800
CTCCGATGAC	CGATATGGTG	GTACTTGCGA	TCCTGATGGC	TGTGACTGGA	850
ACCCATACCG	CTTGGGCAAC	ACCAGCTTCT	ACGGCCCCGG	CTCGAGCTTC	900
ACCTTCGACA	CCACCAAGAA	GATGACCGTC	GTCACCCAGT	TCGCTACTTC	950
GGGTGCCATC	AACCGATATC	ATGTCCAGAA	TGGCGTCACT	TTCCAGCAGC	1000
CCAACGCCGA	GCTCGGTAGC	TACTCTGGCA	ACACGCTCAA	CGATGCTTAC	1050
TGCGCAGCTG	AAGAGGCGGA	ATTGCGCGGA	TCATCTTTCT	CAGACAAGGG	1100
TGGCCTTACC	CAATTCAAGC	AGGCTACTTC	AGGCGGCATG	GTCTTGGTAA	1150
TGAGCCTGTG	GGATGACGTG	AGTTCATGGA	TAGCATTGAC	ATTGTTCGAGA	1200
GAACCATAGC	CGCTGACCGA	GACACAACAG	TACTACGCCA	ACATGCTGTG	1250
GCTGGACTCC	ATCTACCCGA	CGAACGAGAC	CTCCTCTACC	CCCGGTGCCG	1300
CGCGCGGAAG	CTGCTCTACC	AGCTCCGGTG	TCCCTGCCCCA	GCTCGAGTCT	1350
CAGTCTACCA	ACGCCAAGGT	CGTCTTCTCC	AACATCAAGT	TCGGACCCAT	1400
TGGCAGCACT	GGTAACTCCA	GCGGCGGAAA	CCCCCGGGC	GGAGGAAACC	1450
CCCCCGGCAC	CACCACCACC	CGACGCCCAG	CTACCACCAC	CGGAAGCTCT	1500
CCCGGACCTA	CTCAGACACA	CTATGGCCAG	TGCGGTGGAA	TTGGGTACTC	1550
GGGCCCCACG	GTCTGCGCCA	GCGGCAGCAC	ATGCCAGG		1588

Figure 7: *T. konilangbra* genomic DNA.

MYRKLAVITA FLATARA

17

Figure 8A: T. konilangbra signal sequence.

QSACTIQAET	HPPLTWQKCS	SGGSCTSQTG	SVVIDANWRW	THATNSTTNC	50
YDGNTWSSSL	CPDNESCAKN	CCLDGAAYAS	TYGVTTTSADS	LSIGFVTQSQ	100
QKNVGARLYL	MASDTTYQEF	TLLGNEFSFD	VDVSQLPCGL	NGALYFVSMD	150
ADGGVSKYPS	NTAGAKYGTG	YCDSQCPRDL	KFINGEANVE	GWEPASNNAN	200
TGIGGHGSCC	SEMDIWEANS	ISEALTPHPC	TTVGQAICDG	DGCGGTYSDD	250
RYGGTCDPDG	CDWNPYRLGN	TSXYGPGSSF	TLDTTKKMTV	VTQFATSGAI	300
NRYYVQNGVT	FQQPNAELGS	YSGNTLNDAY	CAAEAEFGG	SSFSDKGGLT	350
QFKQATSGGM	VLVMSLWDDY	YANMLWLDSI	YPTNETSSTP	GAARGSCSTS	400
SGVPAQLESQ	STNAKVVFNS	IKFGPIGSTG	NSSGGNPPGG	GNPPGTTTTR	450
RPATTTGSSP	GPTQTHYGQC	GGIGYSGPTV	CASGSTCQVL	NPYYSQCL	498

Figure 8B: T. konilangbra mature amino acid sequence.

TCGGCCTGCA	CCCTCCAGAC	GGAAACTCAC	CCGCCTCTGA	CATGGCAGAA	50
ATGCTCATCT	GGTGGCACTT	GCACCCAACA	GACGGGCTCC	GTGGTCATCG	100
ACGCGAACTG	GCGCTGGACT	CACGCTACGA	ACAGCAGCAC	GAAGTGCTAC	150
GACGGTAACA	CTTGGAGCTC	AACCTTGTGC	CCTGACAATG	AGACTTGCGC	200
GAAGAACTGC	TGCTTGGATG	GTGCCGCCTA	CGCGTCGACG	TACGGAGTCA	250
CCACGAGCGC	TGACAGCCTC	TCCATTGGCT	TCGTCACTCA	GTCTGCGCAG	300
AAGAATGTCG	GCGCCCGTCT	CTACTTGATG	GCGAGTGACA	CGACCTACCA	350
AGAATTTACC	CTGCTTGGCA	ACGAGTTCTC	CTTCGATGTT	GATGTTTCCC	400
AGCTGCCGTA	AGTGGCCAAC	TACACCCCTT	GACGGTATCC	TCTCATTGGT	450
TCCCAGCTGA	CTCGCGAAAT	TAAGATGTGG	CTTGAACGGA	GCTCTTTACT	500
TCGTGTCCAT	GGACGCGGAT	GGTGGCGTGA	GCAAGTATCC	CACAAACACT	550
GCCGGCGCCA	AGTACGGCAC	GGGTTACTGT	GACAGCCAGT	GCCCTCGTGA	600
TCTCAAGTTC	ATCAACGGCG	AGGCCAACGT	TGAGGGCTGG	GAGCCGTTCT	650
CCAACAACGC	CAACACGGGC	ATTGGCGGAC	ATGGAAGCTG	CTGCTCTGAG	700
ATGGATATCT	GGGAGGCCAA	CTCCATCTCT	GAGGCTCTTA	CTCCTCATCC	750
TTGCACGACC	GTCGGGCAGG	AAATTGCGA	TGGTGACTCC	TGCGGCGGAA	800
CCTACTCCGG	TGATCGATAT	GGCGGTACTT	GCGATCCTGA	TGGCTGCGAT	850
TGGAACCCAT	ACCGCTTGGG	CAACACCAGC	TTCTACGGGC	CCGGCTCAAG	900
CTTCGCTCTT	GATACCACCA	AGAAGTTGAC	CGTTGTCACC	CAGTTCGAGA	950
CTTCGGGCGC	TATCAACCGG	TACTACGTCC	AGAATGGCGT	CACTTTCCAG	1000
CAGCCCAACG	CCGAGCTCGG	TAGTTACTCT	GGCAACTCGC	TCGACGATGA	1050
CTACTGCGCG	GCTGAAGAGG	CGGAGTTTGG	TGGCTCTTCT	TTCTCGGACA	1100
AGGGCGGCCT	TACTCAATTC	AAAAAGGCTA	CTTCCGGTGG	CATGGTCTTG	1150
GTCATGAGCT	TGTGGGATGA	TGTGAGTTCA	TGAATAGCAT	TCAAACAGTC	1200
AACAGAATAA	CAGCAGCTGA	CTGAGACACA	ATAGTACTAC	GCCAACATGC	1250
TGTGGCTGGA	CTCCACCTAC	CCGACGAACG	AGACCTCTTC	CACCCCGGT	1300
GCCGTGCGCG	GAAGCTGCTC	CACCAGCTCC	GGTGTCCCTG	CTCAGCTTGA	1350
GTCCCAGTCT	TCCAACGCCA	AGGTCGTATA	CTCCAACATC	AAGTTCGGCC	1400
CTATCGGCAG	CACCGGCAAC	TCCAGCGGCG	GTAGCCCTCC	CGGCGGAGGA	1450
AACCCTCCCCG	GTACCACGAC	CACCCGCCGC	CCAGCTACCT	CCACTGGAAG	1500
CTCTCCCCGGC	CCTACTCAGA	CGCACTATGG	CCAGTGCGGT	GGTATTGGGT	1550
ACTCGGGCCC	CACGGTCTGC	GCGAGTGGCA	GCACTTGCCA	GG	1592S

Figure 9: *T. pseudokonigii* genomic DNA sequence.

MYRKLAVITA FLATARA

17

Figure 10A: T. pseudokoningii signal sequence.

QSACTLQTET	HPPLTWQKCS	SGGTCTQQTG	SVVIDANNRW	THATNSSTNC	50
YDGNTWSSTL	CPDNETCAKN	CCLDGAAYAS	TYGVTTADS	LSIGFVTQSA	100
QKNVGARLYL	MASDTTYQEF	TLLGNEFSFD	VDVSQLPCGL	NGALYFVSMD	150
ADGGVSKYPT	NTAGAKYGTG	YCDSQCPRDL	KFINGEANVE	GWEPPFSNNAN	200
TGIGGGHSCC	SEMDIWEANS	ISEALTPHPC	TTVGQEICDG	DSCGGTYSGD	250
RYGGTCDPDG	CDWNPYRLGN	TSFYGPGSSF	ALDTTKKLTV	VTQFETSGAI	300
NRYYVQNGVT	FQQPNAELGS	YSGNSLDDDY	CAAEAEFGG	SSFSDKGGLT	350
QFKKATSGGM	VLVMSLWDDY	YANMLWLDST	YPTNETSSTP	GAVRGSCSTS	400
SGVPAQLESQ	SSNAKVVSNN	IKFGPIGSTG	NSSGGSPPGG	GNPPGTTTTR	450
RPATSTGSSP	GPTQTHYGQC	GGIGYSGPTV	CASGSTCQVL	NPYYSQCL	498

Figure 10B: T. pseudokoningii mature amino acid sequence.

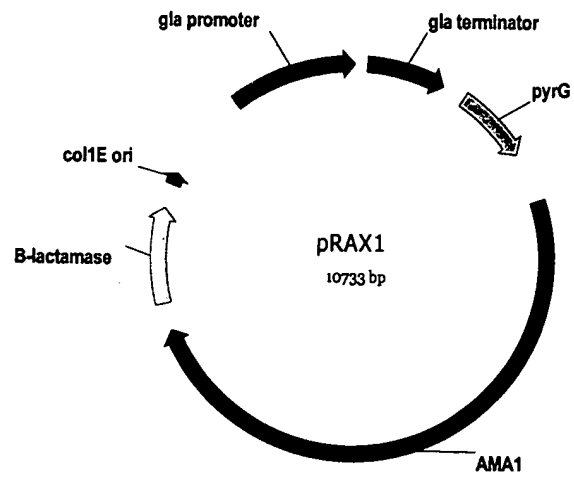


Figure 11: pRAX1

Figure 12: Destination vector pRAXdes2 for expression in *A. niger*

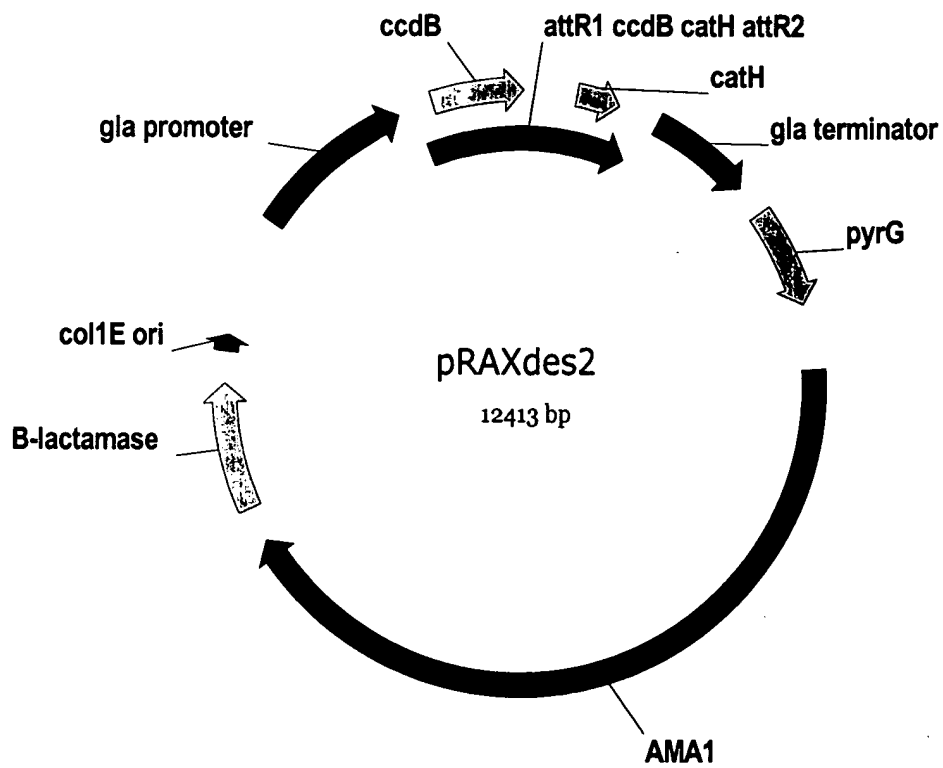


Figure 13: Replicative expression pRAXdesCBH1 vector of CBH1 genes under the control of the glucoamylase promotor.

